

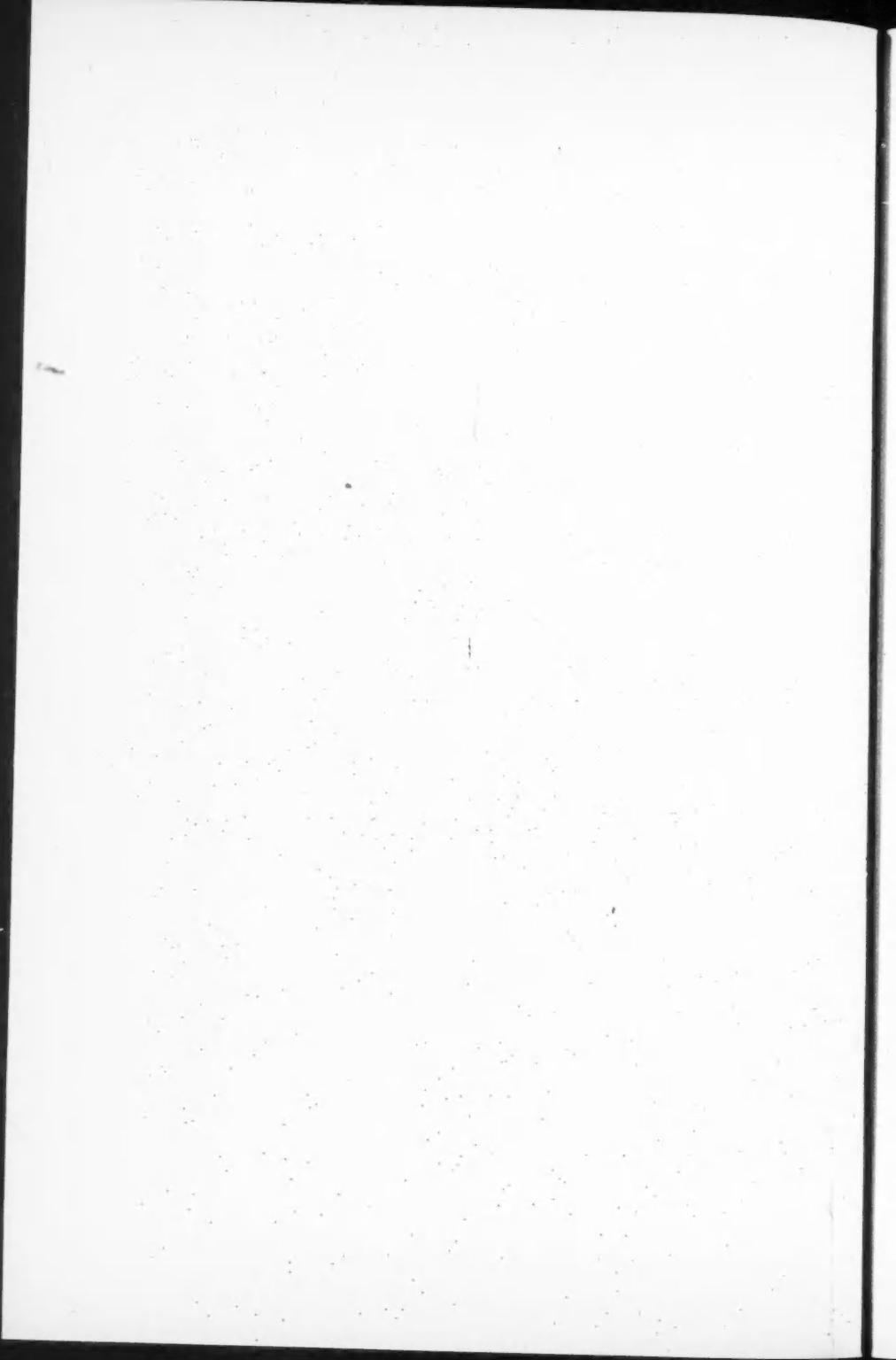
E DUCATIONAL TESTING

by

Helen B. Shaffer

| | Page |
|--|------------|
| TESTING IN MODERN SCHOOL PROGRAMS | 935 |
| Concern Over Extensive Dependence on Tests | 935 |
| Varieties and Purposes of Standardized Tests | 936 |
| Need of Testing to Prevent Waste of Talent | 937 |
| Federal Aid for Testing in New Education Act | 939 |
| EVOLUTION OF EXAMINATION TECHNIQUES | 940 |
| Experiments in Measuring Mental Capacity | 941 |
| Wide Acceptance of Standardized Testing | 942 |
| Development of College Board Examinations | 944 |
| TRENDS AND ISSUES IN SCHOOL TESTING | 946 |
| Limitations of Standardized Aptitude Tests | 947 |
| Controversy Over College Admission Tests | 948 |
| Test Results in Public and Private Schools | 949 |
| Unmeasurable Factors in Testing Capability | 951 |

**Dec. 17
1958**



EDUCATIONAL TESTING

THE PRACTICE of putting American elementary and high school students through uniform, standardized tests, designed to measure accurately a wide range of competencies, was given a strong boost last September when the National Defense Education Act authorized federal aid to expand and strengthen testing programs. Scores obtained on the standardized tests carry considerable weight in determining which high school pupils will take college preparatory courses and which graduates will be admitted to college. The tests are being used increasingly in the lower grades to help evaluate pupil ability and progress and, in some cases, to help determine whether a child shall be assigned to a fast or a slow learning group.

Half a century ago, most of the examinations given to school pupils were made up by the classroom teacher and were designed to test the students' knowledge of subjects they had been taught. Today's pupils still must take examinations of that sort, but they are repeatedly subjected also to tests prepared by specialized agencies for the purpose of gauging general or particular aptitudes or abilities not necessarily directly related to daily school experience.

The general public, and especially parents of school-age children, are naturally concerned that the standardized tests give fair evaluations. Competition for admission to college, particularly to the high-prestige institutions, is growing keener. The increasing availability of college scholarships puts still another premium on high test scores. And standardized tests may be a deciding factor in placing a younger child on an educational track which will broaden or limit his later academic or vocational opportunities.

Educators and psychologists in the testing field try to allay parental doubts by pointing out that scores obtained in the tests are only one of numerous criteria used to judge

individual capacities and potentialities. But the phenomenal growth of standardized testing and the important stakes involved for the individual have led to some questioning of the process. Public acceptance has not been helped by the fact that the profession of testing has developed a jargon unintelligible to the average person, and that scoring is arrived at by mathematical methods beyond common understanding. The practice of most school authorities of withholding the intelligence scores of pupils also has tended to nourish suspicion among parents.

VARIETIES AND PURPOSES OF STANDARDIZED TESTS

Standardized tests used in wide-area testing programs take many forms. Intelligence tests of various kinds are used to measure native capacity to learn. Separate tests usually are given to gauge verbal intelligence, ability to deal with number problems, and capacity to handle abstract ideas. Included in this battery¹ are diverse aptitude tests which attempt to give a more precise evaluation of special abilities in narrowed-down fields of activity.

Achievement tests are more akin to traditional schoolroom examinations in that they measure acquired learning in specific fields. They differ, however, in that they are not aimed so much to test an individual's recollection of material covered in class as to measure his over-all grasp of a subject. An effort is made to construct the tests in such a way that differences among schools in the content of particular courses, and advantages that might be gained from last-minute cramming, will not affect the scores. A variant of the achievement test is one which tests not only knowledge in a particular area but also ability to apply that knowledge to new problems or situations.

Intelligence and achievement tests are the standardized tests most widely used. Other types coming more and more into use are tests to determine personality traits, attitudes, and personal interests. Such tests have been available for decades, but they are still largely in the experimental stage. A test to disclose emotional states is sometimes given to pupils with special problems of adjustment.

Mass-program tests are distinguished from ordinary schoolroom examinations primarily by objectivity and

¹ A battery is a group of tests, covering different areas, designed to be given consecutively and with a related scoring system.

Educational Testing

standardization. Objectivity is achieved by selecting questions that will give a fair sampling of each individual's ability or achievement, and by so constructing questions that there can be only one right (and short) answer to each. Scoring them is entirely independent of the scorer's personal judgment and unaffected by any ambiguities which may creep into an answer framed in the student's own words. Many tests are accompanied by a stencil which, when placed over the test sheet, instantly exposes the correct answers. Tests used in large-scale programs are put through automatic grading machines.

Standardized tests are tests which have been given experimentally to an appropriate sampling of students, whose scores have been processed to arrive at a norm of performance. Pupils do not pass or fail these tests; they simply attain ratings which are compared with the norm. If the norm is derived from the performance of a nationwide or state-wide sample of pupils, an individual's score will differ from what it would be if the norm were derived from the performance of a selected group.

Methods of scoring the tests are highly complex, and the resulting marks not easy to understand. Greatest interest probably attaches to the "I.Q." yielded by intelligence tests. These tests ordinarily are scored according to age. If the child gets a score which is the average obtained by children of a certain age, he is said to have a mental age of that figure. If his mental age is the same as his chronological age, he has an I.Q.² of 100. If his mental age is above his age in years, his I.Q. is higher than 100; if it falls below, his I.Q. is under 100. An I.Q. of 90-110 is usually regarded as representing average mentality; most public schools can accommodate an I.Q. down to 80 or even 75. Gifted children have I.Q.s of 130 or above.

NEED OF TESTING TO PREVENT WASTE OF TALENT

Many educators consider standardized tests a necessity in a mass educational system. They feel that some selective factor is needed to avoid holding all children to an average standard of achievement, which would frustrate the dull and do injustice to the bright. The selection process should be as objective as possible out of fairness to the students

² Intelligence Quotient; that is, the mental age divided by the chronological age and multiplied by 100.

and in recognition of the country's manpower requirements. The nation's need of a highly trained population, each person educated to the utmost of his capacity in the field of his greatest gift, is what has given urgency to demands for more educational differentiation.

President Eisenhower emphasized this point when he submitted proposals for education legislation to Congress last Jan. 27. Noting that many gifted high school graduates fail to go to college,³ the President said there was an "emergency need" for a federal program for "reducing this waste of talent." He added: "Much of this waste could be avoided if the aptitudes of these young people were identified and they were encouraged toward the fullest development of their abilities."

The President followed current educational thought when he recommended a program to encourage development of guidance and counseling programs in the schools. The counselor's job is to help each child mark out the educational path that best fits his talents and interests. That path is not likely to be found unless the guidance counselor has some means of helping the student make intelligent decisions. Traditionally, decisions on the courses to take were made on the basis of report cards, teachers' opinions and other elements of the accumulated school record. The modern educator has added a new and impressive instrument for individual evaluation: the standardized objective test.

Testifying on Jan. 8, 1958, before the House Education Committee, U.S. Commissioner of Education Lawrence G. Derthick said:

The comment is often made that every teacher knows the good students in the class. . . . This . . . comment . . . is all too often misleading. . . . It is not that simple. Getting high grades in a given course is not a reliable indicator of ability. . . . To identify effectively the talents . . . of young people . . . scientific testing procedures must be used. Only after such reliable identification do the counselors and teachers have tools necessary for proper guidance and counseling.

The commissioner pointed out that a conscientious student may get higher grades than a brilliant one who has become inattentive for lack of stimulus to effort. The elective system in high schools allows many talented students

³ According to the U.S. Office of Education, one-half of the high school graduates in the upper half of their class do not go full-time to college and one-fifth don't go at all.

Educational Testing

to drift into easy courses, neglecting subjects necessary to prepare them for college work in needed specialties. Competition for good grades, necessary for admission to high-prestige colleges, causes some students to prefer an A in an easy subject to a lower mark in a difficult subject like physics or trigonometry.

Derthick said the administration bill, then pending, was designed "to stimulate the development of a program of state and local action which will provide the essential professional and administrative leadership in the state departments of education, adequate state supervisory services, and expansion and improvement of testing, counseling, and guidance programs in local schools." Although 41 states had guidance and counseling services, they employed a total of only 63 persons. A recent survey, he said, showed staff shortages, inadequate preparation of staff members, and meager financial support of the programs.

Henry Chauncey, president of the Educational Testing Service—a leading producer of standardized tests—told the committee that student ability should be ascertained in the eighth or ninth grade at the latest, because critical educational decisions had to be made at that stage "with or without the information for making them wisely." Grades alone were not sufficient because marking systems varied greatly among different schools and teachers, and because some pupils might not have demonstrated in class what they could do. "Aptitude tests," he said, "provide a comparable set of observations of pupils that are unaffected by these influences."⁴

FEDERAL AID FOR TESTING IN NEW EDUCATION ACT

The National Defense Education Act authorized \$887 million of federal aid for the nation's schools. Of that total, \$60 million was earmarked for a four-year program to build up guidance, counseling and testing programs in schools and colleges.⁵ Many educators believe that, despite the relatively small percentage of the total going to testing and guidance, this section of the act in the end will have the most telling influence. This is because testing strikes at the heart of every branch of learning, affects funda-

⁴ Testimony, House Education Committee, Feb. 26, 1958.

⁵ Although the terms "guidance" and "counseling" are used almost interchangeably, guidance refers primarily to assistance in mapping a student's educational program and counseling to advice on any kind of school or personal problem.

Editorial Research Reports

mental policy for the educational treatment of all children, and influences course content and teaching methods.

The federal program has been launched in two areas, one for higher education, the other for lower schools. An initial appropriation of \$2 million was made available for the higher education program. Notices were sent to registrars of all colleges and universities, public and private (about 1,900, including junior colleges), inviting them to apply for assistance in establishing institutes for training guidance personnel.

To initiate the lower-school part of the program, regulations were issued in November outlining requirements to be met by the states to qualify for federal aid for school guidance and testing programs. Congress appropriated \$5.4 million for this activity in the present fiscal year.⁶ Some of the money will be used directly for purchase of existing standardized tests or for compiling new tests for use on a state-wide basis. It is expected that some state plans will be approved and ready for application before the end of the current semester.

Evolution of Examination Techniques

MODERN educational testing is an outgrowth of early experiments by psychologists in the realm of mental measurement—experiments which coincided with a movement among educators to put testing of student achievement on a scientific basis. Much of today's testing is predicated on what now seems the obvious fact that individuals differ considerably in learning capacity and in other traits bearing on scholastic achievement. An English scientist, Francis Galton (1822-1911), was the first to demonstrate individual differences in mental ability by means of tests and statistical procedures. His findings impressed educators who had tended to blame the poor learner's inadequacies on laziness.

Educators at the turn of the century were intensely interested in the suggestion that mental capacity could be

⁶The grants this year require no matching; beginning next year, states must put up an amount at least equal to the federal grant.

Educational Testing

mathematically measured, but they lacked an acceptable instrument for measuring it. A French psychologist, Alfred Binet, was the first to create an effective measuring device. Binet and an assistant, Theodore Simon, developed a series of graded tests to determine the ability of children to meet the demands of primary education. Among the things tested was whether a child knew left from right, could follow simple directions, and could recall series of numbers.

EXPERIMENTS IN MEASURING MENTAL CAPACITY

The Binet test, first published in 1905 and revised in 1908 and 1911, attracted world-wide attention and launched a flurry of experimentation. A recent textbook observed that Binet's "method and materials for the measurement of intelligence form the basis of the general approach today."⁷

Binet's tests were given individually and required the services of a trained examiner, who had to spend at least an hour with each child. The first effective group test of mental ability that could be given simultaneously to large numbers was developed by the U.S. War Department during World War I. A group of psychologists, headed by Edward L. Thorndike, produced the famous Army Alpha, which was used to classify more than a million draftees by intelligence levels.

The Army Alpha was a verbal test designed to be taken only by persons with at least a sixth-grade education. Another test, Army Beta, was devised for illiterates and draftees who spoke only a foreign language. A personality test was developed for identification and study of suspected neurotics among draftees, and some progress was made in working out aptitude tests to identify suitable candidates for specific kinds of duty.

After the war, the Army tests were made available for civilian use. Thousands of high school and college students took the Army Alpha, and psychologists and educators set out to develop tests to evaluate nearly every aspect of human ability. Many educators subscribed to the Thorndike dictum that "Anything that exists . . . exists in some quantity; and anything that exists in some quantity is capable of being measured." Question construction tech-

⁷ Victor H. Noll, *Introduction to Educational Measurement* (1958), p. 22.

Editorial Research Reports

niques and statistical methods of scoring, first developed for the mental measurement tests, were adapted to tests for evaluating student achievement. Teachers in large numbers discarded the traditional essay-type question in favor of the true-or-false, multiple-choice, or other short-answer type. The belief that a student's progress should be measured against his native capacity warred with the belief that it should be measured against the norm of all children in his grade.

A tremendous literature on educational testing began to accumulate. Professional societies and journals were established; many large universities instituted testing research activities;⁸ and commercial enterprises were launched to publish and distribute standardized tests. By 1928 more than 1,100 different standardized tests had been compiled, and sales of copies had hit an annual total of between 20 million and 30 million. At the same time, much research was being directed to refining question selection and construction, grading procedures, and statistical methods of evaluating scores; also to double-checking the reliability of existing tests.

WIDE ACCEPTANCE OF STANDARDIZED TESTING

World War II put another powerful spur to the testing movement. All branches of the armed services used standardized tests as an aid to classifying and assigning personnel. They also conducted research on the nature of human capabilities and developed new forms of tests and rating scales. According to one authority, the tests developed by the armed forces did "more to improve reliability of testing for student guidance purposes than any other single thing."⁹

Cooperation between colleges and universities and the military agencies, beginning in World War II, gave many educational institutions their first intensive experience in the development and use of standardized tests. Off-duty educational programs sponsored by military agencies employed such tests on a wide scale. The General Educational Development test, for example, was taken by thousands of students in service to establish their educational

⁸ The first university testing center was established in 1913 at the University of Oklahoma, but most of the present centers did not come into existence until after World War I.

⁹ Ralph F. Berdie, testimony, House Education Committee, Oct. 28, 1967.

Educational Testing

status; many local school authorities accepted certain ratings on this test as the equivalent of high school graduation, as did some colleges and many employers.

Standardized testing is firmly established today. Arthur E. Traxler, executive director of the Educational Records Bureau (a test-service agency with 720 school members), estimated last spring that the country's educational institutions would use 108 million copies of standardized tests in the school year 1957-58.¹⁰ More than 5,000 separate tests have been developed since modern testing started, and more than 1,000 are now given. According to Traxler, 25 or 30 "titles" account for the bulk of the tests used. At least three-fourths of all tests are given in the elementary grades. This indicates that the schools are moving toward regular and repeated testing which will build up for their pupils a cumulative record of ability and achievement. Twenty-five states have state-wide testing programs.

There are approximately 20 test publishers in the United States,¹¹ all required to observe standards laid down by the American Psychological Association, the American Educational Research Association, and the National Council on Measurements Used in Education. "Test preparation and publishing," Traxler told a recent conference on testing, "is now a rather well-defined science." Every item on a test is subjected to a rigorous procedure to determine its suitability for measuring the factor in question.

It was inevitable that the testing movement should have great influence on college admission standards. In colonial times the only scholastic requirement for matriculation was command of Latin and Greek. Applicants for admission to Harvard in the early 19th century took an all-day (6 a.m. to 6 p.m., with half-hour off for lunch) oral test in Greek, Latin and arithmetic. Written examinations were not required until mid-century.

By the end of the last century each college had its own set of admission standards, which varied in considerable degree according to the particular Greek or Latin works on which applicants were quizzed. The wide range of the

¹⁰ Fifty-five per cent of the estimated total was composed of achievement tests, 35 per cent of ability or intelligence tests, and 10 per cent of tests for measuring interest or personality.

¹¹ Leading publishers include the Educational Testing Service, the Psychological Corporation, World Book Co., California Test Bureau, Science Research Associates, Houghton Mifflin Co.

Editorial Research Reports

examinations posed a difficult problem for secondary schools attempting to prepare students for admission to college. Acceptance by some colleges in other states of New York State pupils who had passed the New York Board of Regents examinations introduced a small measure of uniformity into a situation of general disorder. However, the Regents' examinations for certifying college eligibles were criticized by educators of that day as erratic if not unfair.

Two prominent figures in American education, Nicholas Murray Butler,¹² president of Columbia University, and Charles W. Eliot, president of Harvard University, initiated a movement to bring order into the admissions picture. Their efforts led to the creation in 1900 of the College Entrance Examination Board, a cooperative undertaking of colleges and secondary schools to draw up and administer uniform college admission tests. The first examinations were given in 1901 to 973 candidates; their 8,000 examination papers were graded by a corps of 39 "readers" instructed to follow a method prescribed by the board.

It was years before more than a handful of eastern colleges relinquished the prerogative of administering their own admissions tests, although some of the institutions which continued to give their own examinations would accept candidates who had passed the "college boards." The new plan eventually took firm hold, and its recent growth has been striking. Around 300 institutions now use the examinations either as a screening device or as an aid to student selection and placement. Although many state universities are required to accept all graduates of accredited high schools within the state, several of them (Florida and Texas, for example) have instituted state-wide testing to limit enrollments. According to Frank Bowles, president of the C.E.E.B., approximately one-half of the 1,000 four-year, degree-granting colleges and universities in the country require applicants to take admission examinations; he predicted recently that within five or ten years all of these institutions would do so.

DEVELOPMENT OF COLLEGE BOARD EXAMINATIONS

College Board examinations in the early days followed traditional testing procedures. Questions were addressed

¹² Butler had been admitted to Columbia only conditionally, in 1878, because he failed to recite in order the names in Latin and English of all the capes and rivers of Europe.

Educational Testing

to specific course subjects, and they often reflected the scholarly prejudices of the academic personnel engaged in test-making. Questions requiring lengthy answers resulted in considerable unevenness in grading.

By the second decade of the century, educators had become enamored of the concept of the "comprehensive examination"; that is, a test intended to disclose not only the extent of an applicant's fund of factual information but also his "ability to reason independently and to compare and correlate the material of a broad field of study."¹³ The comprehensive type of test was first given by the College Board in 1916. During the post-World War I years, the board was influenced by the crusade for intelligence testing as a better predictor of college performance than subject content tests. The first scholastic aptitude test, given in 1926, was split three years later into two parts: one to measure verbal intelligence, the other to test fitness for advanced study of science and mathematics.

The board offered two series of tests for college admission at that time. One, given in April, was a battery of aptitude and achievement tests subject to automatic scoring. The other tests, given over a week's period in June, were of the conventional type requiring essay answers and covering various subject fields. Member colleges could employ either or both in qualifying applicants; the trend was toward the aptitude test as the more useful in disclosing fitness for college.

When the United States entered World War II in 1941, the essay examination series was dropped and was not thereafter resumed. College Board examinations today include both aptitude and achievement tests of the objective, standardized type. Member colleges decide which tests they will require of their applicants and make their own evaluations of the ratings.

The tests now are drawn up, given at numerous centers around the country and abroad, and automatically scored by the Educational Testing Service of Princeton, N. J. E.T.S. was founded in 1947 by the College Entrance Examination Board and the American Council on Education to take over and merge their testing activities. Grants from the Carnegie Corporation and the two parent bodies

¹³ A. Lawrence Lowell, "The Art of Examination," *Atlantic Monthly*, January 1926.

launched E.T.S. on a non-profit program of test-making and research. It currently supplies not only the tests given by the College Board but also tests for admission to schools of law, medicine and other professions, qualifying tests for applicants to the Department of State's Foreign Service, and tests used by several extensive scholarship plans, including the National Merit Scholarship Program.

Trends and Issues in School Testing

A HALF-CENTURY of experimentation has failed to bring agreement on the effectiveness of objective test techniques or on the comparative merits of types of tests and rating scales now in use. The more those who make and study the tests learn about measurement of innate and acquired abilities, not to mention other less tangible human traits and attitudes, the more complex the problem of testing seems to become.

The question that most troubles parents is whether tests in use provide a fair method of differentiating among students. Even the strongest advocates of modern scientific testing, while contending that the system works well for the majority—especially for those at the upper and lower extremities of the rating scale—admit that there are always occasional individuals whose peculiar abilities, or potential abilities, may not be fully disclosed by a particular test given at a particular time.

Detlov W. Bronk, president of the National Academy of Sciences-National Research Council, cited an instance of a test misfire in testimony before the Senate Labor and Public Welfare Committee last Jan. 21. Bronk said his son had been turned down some years ago by a leading preparatory school after receiving a low score on the school's admission test; the school advised the father that the boy did "not have the intellectual capacity for higher education." However, Bronk's son was admitted to another school and was graduated second in a class of 176, was "top of his class" at Princeton, and won a Rhodes Scholarship.

The value placed on standardized tests in the United States puzzled Anatolli A. Smirnov, Russian psychologist,

Educational Testing

who inspected many American schools in November as a member of a Soviet team of educators. In an interview Dec. 12 he said: "I and my colleagues have worked with the tests for 22 years. We have concluded that they are virtually worthless. They are not a fair measure either of mental ability or of achievement." The Russian student's placement and advancement depends on what Smirnov called "extra-school" activities—participation in technical clubs, academic competitions, and the like.

LIMITATIONS OF STANDARDIZED APTITUDE TESTS

Challenged on the reliability of the standardized tests, the president of the Educational Testing Service, Henry Chauncey, told the House Education Committee that the merits of the tests as forecasters of future success in college and in the professions had been studied thousands of times, and the studies showed "on the average . . . a very substantial degree of relationship." He added: "This is not to say that there aren't always some individuals who do much better than you would expect, and some individuals who don't do as well as you would expect," but this was also the experience when sole dependence was placed on an individual's school record. Young people from "underprivileged environments," for example, could not be expected to do as well on the tests, relative to their native capacity, as "more culturally favored pupils."

The standardized tests have recognized limitations. "We have no particular evidence that they measure potential creativity, original thinking, inventiveness," Chauncey said at a conference last winter. "They certainly will not single out for us the individual who will discover new intellectual territory as distinct from the other individuals who will settle and cultivate that territory." The most that can be expected is that the tests will help "identify the larger number of students who are in the score ranges from which creative scientists, engineers, philosophers, historians, economists, psychologists, jurists, educators are most likely to emerge."¹⁴

Follow-up studies of students who have taken standardized scholastic aptitude tests at age 14 indicate that of those who score in the top 20 per cent, 45 per cent will do

¹⁴ Henry Chauncey, "How Tests Help Us Identify the Academically Talented," address before Conference on the Academically Talented sponsored by National Education Association, Feb. 6-8, 1968.

Editorial Research Reports

honor work in college, 52 per cent will do satisfactory work, and about 3 per cent will fail. The figures are reversed for those scoring in the bottom 20 per cent. In the middle group, constituting 60 per cent of the total, honor work will be done by 17 per cent, satisfactory work by 66 per cent, and the remaining 17 per cent will fail. Another study, covering 281 men listed in *Who's Who in America* and in *American Men of Science* who had taken the College Board's scholastic aptitude tests in their youth, showed high correlation between test scores and post-college achievement.

CONTROVERSY OVER COLLEGE ADMISSION TESTS

The big debate over educational testing results from recognition of the fact that no battery of tests, within reasonable limits of testing time, can measure with certitude all aspects of an individual's qualifications for advanced study. Some educators think that achievement tests should be scrapped in favor of the scholastic aptitude battery. The advantages would be less duplication of testing, possible discovery of obscure talent, and removal of any undue influence on high school curriculums.

A specialist attending a recent conference on testing said that "The growing prevalence of national and state-wide programs embodies a real danger that individual differences among institutions of higher education will be overlooked." The result would be that "The advantages of uniform testing programs may be purchased at the excessive price of ignoring one of the greatest strengths of our educational system—the variety of functions performed by our colleges and universities." The natural tendency of teachers to conduct classes with college admission tests in mind, and the tendency to evaluate teaching on the basis of whether "students do well on some esteemed achievement tests," were deplored.¹⁵

Another specialist took the opposite view by advocating still greater uniformity of testing. High school schedules were being disrupted, he asserted, by too many different test programs and too many kinds of tests. He urged development of a single multiple-purpose test to serve numerous purposes: college admission, scholarship awards,

¹⁵ Alexander G. Wesman, Invitational Conference on Testing Problems, New York City, Nov. 1, 1958.

Educational Testing

admission to professional schools, student guidance, and school evaluation.

According to this view, there has been too much unrealistic reverence for pure mental ability, which actually does not exist apart from the acquisition of knowledge or skills for the exercise of ability. A proper multiple-use test would consist in large part of "exercises requiring the student to interpret and to evaluate critically the same kinds of reading materials that he will have occasion to read and study in college and, particularly, that will require him to do the same kinds of complex reasoning and problem solving that he will have to do later both in and out of school."¹⁶

Several speakers at the conference on testing expressed dissatisfaction with the tendency to justify certain tests because scores made on them are reliable predictors of future grades in college or of professional and commercial success. Such a criteria may overlook the fact that the educational program itself, or the concept of after-college success, may be at fault. "There are better ways of improving the input to our colleges than by striving to improve the prediction of faulty measures of student success in attaining poorly defined and somewhat questionable goals," one of the speakers declared. "Tests designed only to predict success in current programs of instruction do not adequately measure the characteristics which determine educability."¹⁷

Thinking along these lines has motivated current research aimed to develop tests which will probe complex aspects of intellectual promise. The Educational Testing Service has worked on a "Personality Research Inventory" which yields scores on such factors as insight, tolerance of frustration, tolerance of ambiguity, and other traits which influence an individual's attitude toward learning.

TEST RESULTS IN PUBLIC AND PRIVATE SCHOOLS

Questions have been raised about the norms established for tests, particularly for the tests used for guidance and as instructional aids in the lower schools. In the earlier days of testing, norms were higher because schools interested in cooperating with the testing agencies were gen-

¹⁶ E. F. Lindquist, State University of Iowa, Invitational Conference, Nov. 1, 1958.

¹⁷ Robert L. Ebel, Invitational Conference, Nov. 1, 1958.

Editorial Research Reports

erally of a relatively select group. With the wide expansion of testing, norms more representative of the nation's schools as a whole were developed. The result, however, has been to emphasize the disparity between scores obtained in private schools and those obtained in the great majority of public schools.

An official of the Educational Records Bureau reported at a recent meeting¹⁸ that in the five-year period, 1948-1952, an average of about 72 per cent of non-public school pupils ranked above the public school norm median on a standardized test in American history. During the past six years an average of about 85 per cent of the non-public school pupils tested scored the public school norm median on a new American history test.¹⁹ Growth of prosperous suburban communities has produced a number of public schools in which the scholastic level of the student body is comparable to that in the typical private school. Tests designed for the average public school population offer little challenge to the pupils in these schools and are of limited service in their instructional and guidance programs.

The California Test Bureau, which is one of the major sources of nationally standardized tests, has sought to meet some of the inadequacies of prevailing norms by developing a series of more selectively standardized achievement tests. The series is standardized according to samplings of children who are within six months of the same chronological age, are in the same grade of school, and whose I.Q.s are within the same short-range interval. A pupil taking this test acquires an "Intellectual Status Index," which is supposed to represent his mental ability related to his age-grade scale.

The trend toward more frequent use of tests in the lower grades is expected to overcome some of the shortcomings of the one-shot college admissions test, because it will make an individual's performance over a number of years available for evaluation in the cumulative school record. The Educational Testing Service recently completed the "Sequential Tests of Educational Progress," designed to be given at various times from the fourth grade through the sophomore year of college.

¹⁸ Robert D. North, 23d Educational Conference, New York City, Oct. 30, 1958.

¹⁹ The selectivity of private school enrollment is forcibly demonstrated by an Educational Records Bureau report that only 5 per cent of 12,000 non-public school pupils tested in the autumn of 1957 had I.Q.s below 100, the average for public school pupils.

Educational Testing

UNMEASURABLE FACTORS IN TESTING OF CAPABILITY

As test-makers try to make their measures of human competence more scientific, more and more is heard about the unmeasurable factors in evaluating capabilities. The vogue of the objective test question has never fully downed demands for the type of question that requires a pupil to express himself in his own language, out of his own resources. The Educational Testing Service has grappled for years with the problem of developing a method of automatically testing ability in English composition. But English teachers have insisted that the faculty of self-expression in words must not be overlooked in the effort to make examinations more objective.

With the movement of educational measurement into the more complex areas of human testing, the importance formerly attached to a pupil's innate mental ability, most generally known in the form of his I.Q., has been considerably reduced. Pupils are still scored for an I.Q., but the result is used only as one of many measuring rods to gauge how effectively a teaching program is going over with each child. Most school authorities do not disclose I.Q. ratings to parents, except to report that the rating is average or above or below average. Reluctance to give out the actual score results from a belief that parents generally do not understand that an I.Q. rating serves mainly as an educational tool, and that an average or below-average I.Q. does not necessarily put a permanent limit on what a child can accomplish.

Many voices have been raised to remind testers that some of the most valued qualities of a student, as well as an adult, defy measurement. As a Rockefeller Brothers Fund report on education, last June 22, said: "Decisions based on test scores must be made with the awareness of the imponderables in human behavior. We cannot measure the rare qualities of character that are a necessary ingredient of great performance. We cannot measure aspiration or purpose. We cannot measure courage, vitality, or determination."





